Numer. Math. (2005) 100: 441–456
Digital Object Identifier (DOI) 10.1007/s00211-005-0585-6

Numerische
Mathematik

Rate of convergence of regularization procedures and finite element approximations for the total variation flow

Xiaobing Feng[1], **Markus von Oehsen**[2], **Andreas Prohl**[2]

[1] Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA
e-mail: xfeng@math.utk.edu
[2] Department of Mathematics, ETH Zürich, 8092 Zürich, Switzerland
e-mail: {mvo,apr}@math.ethz.ch

Received August 28, 2003 / Revised version received December 9, 2004
Published online: April 11, 2005 –

Summary. We derive rates of convergence for regularization procedures (characterized by a parameter ε) and finite element approximations of the total variation flow, which arises from image processing, geometric analysis and materials sciences. Practically useful error estimates, which depend on $\frac{1}{\varepsilon}$ only in low polynomial orders, are established for the proposed fully discrete finite element approximations. As a result, scaling laws which relate mesh parameters to the regularization parameter are also obtained. Numerical experiments are provided to validate the theoretical results and show efficiency of the proposed numerical methods.

Mathematics Subject Classification (1991): 65M60, 65N30, 35B25, 35K57

1 Introduction

This paper concerns finite element approximations of the L^2-gradient flow

$$-u_t \in \partial J_\lambda(u), \qquad u(0) = u_0$$

for the total variation functional

$$J_\lambda(u) = |Du|(\Omega) + \frac{\lambda}{2}\int_\Omega |\, u - g\,|^2\, dx \tag{1}$$

Correspondence to: A. Prohl

on a bounded domain $\Omega \subset \mathbf{R}^d$ $(d = 1, 2, 3)$, for given functions $u_0,\ g \in L^2(\Omega)$ and a nonnegative number λ. Where ∂J_λ denotes the subdifferential of the functional J_λ (cf. [9]), $|Du|(\Omega)$ denotes the total variation of the function u defined by (cf. [2])

$$|Du|(\Omega) := \sup\left\{ \int_\Omega -u \operatorname{div} \mathbf{v}\, dx\ ;\ \ \mathbf{v} \in [C_0^1(\Omega)]^d,\ \|\,\mathbf{v}\,\|_{L^\infty} \le 1 \right\},$$

and $BV(\Omega)$ will be used to denote the space of functions of bounded total variation. In the rest of this paper, we refer to the above gradient flow as the $\{$new total variation flow$\}$ or simply, the TV flow.

Since the functional J_λ is convex, by nonlinear semigroup theory existence of a unique strong solution $u \in C\big([0,T]; L^2(\Omega)\big) \cap W^{1,2}_{\mathrm{loc}}\big(0,T; L^2(\Omega)\big)$ and $u(t) \in \mathrm{Dom}(\partial J_\lambda)$ for all $0 < t \le T$ can be shown easily (cf. [5]). However, a rigorous characterization of the subdifferential ∂J_λ is quite involved, and this program has only been carried out recently by F. Andreu, C. Ballester, G. Bellettini, V. Caselles, J. I. Diaz, J. M. Mazón, M. Novaga in a series of papers [3,4,6,8], which, along with other results, are summarized in the recent monograph [5].

The following characterization of ∂J_0 is taken from [5], which crucially uses Anzellotti's (cf. [7]) generalization of Green's formula to measures which are given as products of a bounded vector field and a vector-valued Radon measure.

Theorem 1 *Let $\lambda = 0$, a function $u \in C\big([0,T]; L^2(\Omega)\big)$ is strong semigroup solution of TV flow, if $u \in W^{1,2}_{\mathrm{loc}}\big(0,T; L^2(\Omega)\big) \cap L^1_w\big((0,T); BV(\Omega)\big)$, $u(0) = u_0 \in L^2(\Omega)$, and there exists $\mathbf{z} \in L^\infty\big((0,T) \times \Omega; \mathbf{R}^d\big)$, $\|\,\mathbf{z}\,\|_\infty \le 1$, with $\langle \mathbf{z}(t), \mathbf{n}\rangle = 0$ on $\partial\Omega$, for a.e. $t \in [0,T]$ such that $u_t = \operatorname{div} \mathbf{z}$ in $\mathcal{D}'\big((0,T) \times \Omega\big)$, and*

$$\int_\Omega \big(u(t) - w\big) u_t(t)\, \mathrm{d}x = \int_\Omega \big(\mathbf{z}(t), Dw\big)$$
$$-|\,Du(t)\,|(\Omega) \quad \forall\, w \in L^2(\Omega) \cap BV(\Omega)\,, \quad a.e.\ t \in [0,T]\,.$$

This result provides the basis to verify the following properties of the strong solution of the TV flow (cf. [5]) in the case $\lambda = 0$:

(i) u will be constant of mean value of the initial datum in finite time ($d = 2$; cf. Fig. 1);
(ii) $u(t) \in L^\infty(\Omega)$, $t > 0$, if $u_0 \in L^d(\Omega)$, and in general, no $L^1 - L^2$-regularizing effect for $L^1(\Omega)$-initial data;
(iii) $C^{1,\alpha}$-regularity of level sets $\partial^*[u(t) > \lambda]$ for $u_0 \in L^d(\Omega)$ of decreasing size, i.e., $\frac{d}{dt}\mathcal{H}^{d-1}\big(\partial^*[u(t) > \lambda]\big) \le 0$;
(iv) invariance of supports, provided the curvature of a smooth boundary of a simply connected convex domain is not too big (cf. Figs. 1, 2).

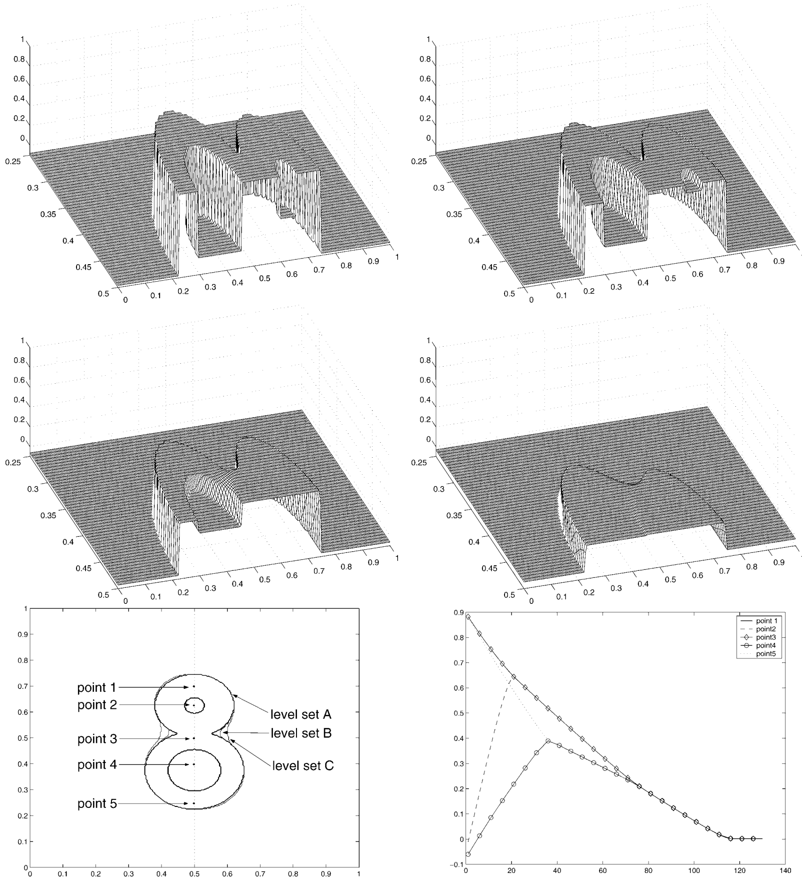

Fig. 1. Evolution of $u_0 = \chi_{\text{torus}}$, using (6)–(8), for $\lambda = 0$: change of support at positive times, different changes of height for points $x \in \Omega$. Plot of evolving heights at five points of Ω: element of torus, inside smaller circle, inside larger circle; $(k, h; \varepsilon) = (10^{-3}, 3.10^{-2}, 10^{-4})$. See also Chapter 4 of [5]

It should be pointed out that the total variation flow was first analyzed by Hardt and Zhou in [18], which studied the gradient flow for a class of linear growth functionals with L^∞ initial data by using the variational inequality techniques introduced in [20]. Lately, Feng and Prohl [15] studied the total variation flow with L^2 initial data using a constructive energy method which can be used to compute the solution of the total variation flow (see the discussion below). On the other hand, the subdifferential ∂J_λ was not characterized in either [18] or [15]. That is, the relationship between the semigroup solution and the weak solutions proved in [18] and [15] was not established.

The best known application of the TV flow arises from image processing for image denoising (see Fig. 3). The well-known noise removal and image restoration model, which was proposed by Rudin, Osher and Fatemi [21], and

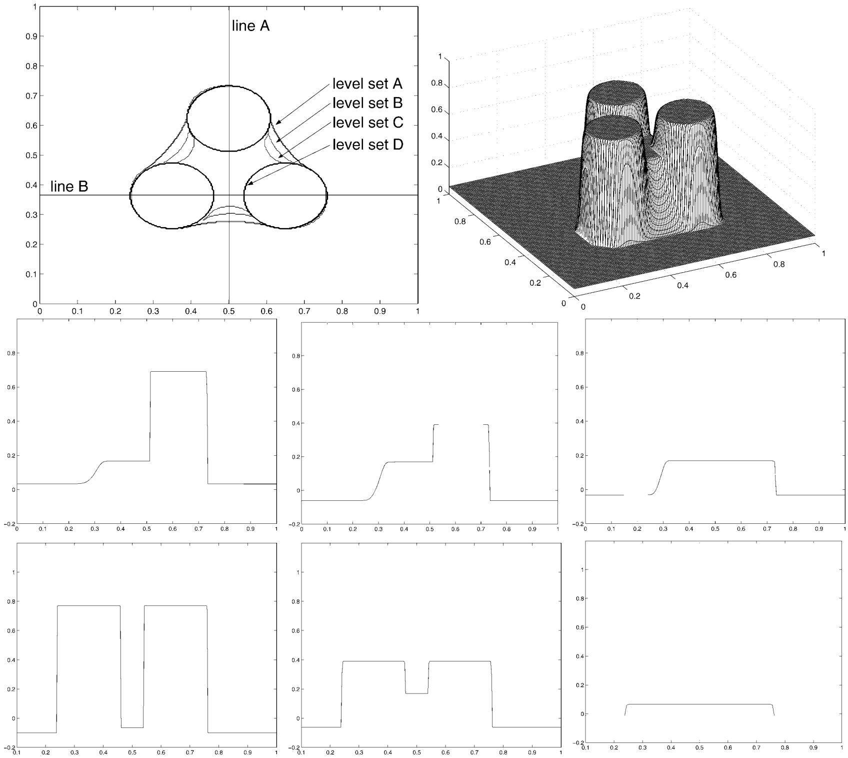

Fig. 2. Evolution of $u_0 = \chi_{3\,\text{balls}}$ for $\lambda = 0$: change of support at positive times (1st row), snapshots of 1D-cross sections of evolution along line A (2nd row), and along line B (3rd row); $(k, h; \varepsilon) = (10^{-3}, 3.10^{-2}, 10^{-4})$. See also Chapter 4 of [5]

analyzed by Acar and Vogel [1], and Chambolle and Lions [13], seeks the minimizer of the functional J_λ as the "best" restored image for a given noisy image g. Solving the minimization problem by using the popular steepest descent method then motivates to consider the above gradient flow. In such an application, the constant λ is known as a Lagrange multiplier which is determined by the original constrained minimization problem (see [13] for a detailed exposition). Additionally, the TV flow also appears in geometric measure theory for studying the evolution of a set with finite perimeter without distortion of the boundary [8] and in materials science for studying the crystalline flow [19].

Formally, the TV flow is described by the following initial boundary value problem

$$(2)\quad \frac{\partial u}{\partial t} = \operatorname{div}\Big(\frac{Du}{|\,Du\,|}\Big) - \lambda\big(u - g\big) \qquad \text{in } \Omega_T \equiv \Omega \times (0, T), \quad T > 0,$$

$$(3)\quad \frac{\partial u}{\partial n} = 0 \qquad \text{on } \partial\Omega_T \equiv \partial\Omega \times (0, T),$$

$$(4)\quad u(\cdot, 0) = u_0(\cdot) \qquad \text{in } \Omega\,,$$

Fig. 3. The initial noisy image and two smoothed samples at times $t = 5 \cdot 10^{-4}, 10^{-3}$ using $f_{\varepsilon,2}$-regularization. $\varepsilon = 10^{-5}$, $k = 10^{-5}$, $h = 2^{-8}$ and $\lambda = 0$ are used in the test

where Du, a vector-valued Radon measure, denotes the distributional gradient of u.

Recently, an algorithm based on the characterization of the subdifferential is proposed by Chambolle [11]; however, so far most existing numerical works deal with a discretization of the gradient flow for the regularized functional

$$(5) \qquad J_{\lambda,\varepsilon}(u) = \int_\Omega f_\varepsilon(|\, Du \,|)\, dx + \frac{\lambda}{2}\int_\Omega |\, u - g \,|^2\, dx\,, \qquad \varepsilon > 0$$

where $f_\varepsilon(z) = \sqrt{z^2 + \varepsilon^2}$, and for any $u \in BV(\Omega)$, the first term on the right hand side of (5) is defined as (cf. [2])

$$\int_\Omega \sqrt{|\, Du \,|^2 + \varepsilon^2}\, dx$$
$$:= \sup\Big\{ \int_\Omega \big[-u \operatorname{div} \mathbf{v} + \varepsilon\sqrt{1 - |\mathbf{v}(x)|^2}\big]\, dx\,;\ \ \mathbf{v} \in [C_0^1(\Omega)]^d,\ \|\, \mathbf{v} \,\|_{L^\infty} \le 1 \Big\}.$$

An explicit characterization of $\partial J_{\lambda,\varepsilon}$ could be obtained by relating the strong semigroup solution to the variational solution of

$$(6) \qquad \frac{\partial u^\varepsilon}{\partial t} = \operatorname{div}\Big(\frac{f_\varepsilon'(|\, Du^\varepsilon \,|)Du^\varepsilon}{|\, Du^\varepsilon \,|}\Big) - \lambda(u^\varepsilon - g) \qquad \text{in } \Omega_T\,,$$

$$(7) \qquad \frac{\partial u^\varepsilon}{\partial n} = 0 \qquad \text{on } \partial\Omega_T\,,$$

$$(8) \qquad u^\varepsilon(\cdot, 0) = u_0(\cdot) \qquad \text{in } \Omega\,,$$

which was carefully analyzed in [15] by exploiting the connection between the flow (6)–(8) and the prescribed mean curvature flow (corresponding to the case $\varepsilon = 1$) (cf. [17,20]). In addition, utilizing the regularity results for the prescribed mean curvature flow [17], i.e., $u_0 \in C^2(\overline{\Omega})$ and $g \in L^\infty\big((0,T); W^{1,\infty}(\Omega)\big)$ implies $u \in W^{1,\infty}(\Omega_T) \cap L^2\big((0,T); H^2(\Omega)\big)$, it was proved in [15] that for each *fixed* ε the implicit Euler-FEM (finite element method) discretization of (6)–(8) converges with *optimal order* to the solution of the problem (6)–(8), provided that the spatial mesh size h and temporal mesh size k satisfy $k = \mathcal{O}(h^2)$. This mesh requirement seems surprising for an implicit scheme, but turns out necessary in the analysis to deal with the singular character of the problem. Although the convergence of the fully discrete finite element solution $U^{\varepsilon,k,h}$ to the solution of the TV flow (2)–(4) was also proved in [15] as both ε and h go to zero and $k = \mathcal{O}(h^2)$, no rate of convergence was established and no precise scaling law which relates h and k to ε was derived there. The goal of this paper is to derive a rate of convergence for the error $u - U^{\varepsilon,k,h}$, where u stands for the solution of the TV flow (2)–(4) and $U^{\varepsilon,k,h}$ denotes the solution of the fully discrete finite element discretization of (6)–(8) proposed in [15].

The remaining of this paper is organized as follows. In Section 2 we establish rates of convergence for various regularization procedures to approximate the TV flow. Possible regularizations include those given in (5) and [13]. Recall that in the case $f_\varepsilon(z) = \sqrt{z^2 + \varepsilon^2}$, it was proved in [15] that u^ε converges to u in L^p for $1 \le p < \frac{d}{d-1}$, but no rate of convergence was given. In Section 3 we revisit the finite element method proposed in [15] for approximating (6)–(8) and (2)–(4), and establish optimal rate of convergence for the finite element scheme for approximating the TV flow (2)–(4). Finally, in Section 4 we provide some numerical experiments, which are especially designed to verify our theoretical results for various regularization strategies, and numerically to find explicit scaling laws which relate regularization and discretization parameters in order to obtain optimal rate of convergence.

This paper is a shortened version of [16] where one can find more details and further helpful comments which could not be included here due to page limitation.

2 Rate of convergence of the regularized flow as $\varepsilon \longrightarrow 0$

From [15], we know that the solution of the regularized flow (6)–(8) converges to the solution of the TV flow (2)–(4) strongly in $L^p(\Omega)$ for $1 \le p < \frac{d}{d-1}$ as $\varepsilon \to 0$. However, it does not tell how fast it converges. We will address the issue in this section by establishing a rate of convergence (in powers of ε). Moreover, we will consider more general regularization procedures by stating some structural assumptions on $J_{\lambda,\varepsilon}$, which cover commonly used regularization procedures; in particular, they include $J_{\lambda,\varepsilon}$ defined in (5) and a modified regularization procedure introduced by Chambolle and Lions in [13].

Theorem 2 *Suppose that $u_0, u_0^\varepsilon \in L^2(\Omega)$ and $g \in L^2(\Omega)$. Let u, u^ε be the weak solutions of* (2)–(4) *and* (6)–(8)*, respectively. Assume there exist two positive constants α, $C_0(T)$ such that $J_{\lambda,\varepsilon}$ satisfies*

$$\int_0^T | J_{\lambda,\varepsilon}(v) - J_\lambda(v) | \, dt \le C_0(T)\varepsilon^\alpha \qquad \forall\, v \in L^1((0,T); BV(\Omega)) \cap L^2(\Omega_T), \tag{9}$$

then, there holds

$$\operatorname*{ess\,sup}_{t\in[0,T]} \| u(t) - u^\varepsilon(t) \|_{L^2(\Omega)} \le \| u_0 - u_0^\varepsilon \|_{L^2(\Omega)} + 2\sqrt{C_0(T)}\, \varepsilon^{\frac{\alpha}{2}} \, . \tag{10}$$

Proof. For any $f \in L^2(\Omega)$ and $0 < \rho \ll 1$, let $(f)_\rho := \mathcal{M}_\rho * f \in C^\infty(\Omega)$ denote its mollification. Here $\mathcal{M}_\rho$ can be chosen as any well-known mollifier.

Let u_ρ and u_ρ^ε denote solutions of TV flow $-u_t \in \partial J_\lambda(u)$ with initial datum $(u_0)_\rho$, and (6)–(8) with initial datum $(u_0^\varepsilon)_\rho$, respectively, then u_ρ and u_ρ^ε satisfy

$$\begin{aligned} &\int_0^s \int_\Omega u_{\rho t}(v - u_\rho)\, dxdt \\ &+ \int_0^s \big[J_\lambda(v) - J_\lambda(u_\rho)\big]\, dt \ge 0 \quad \forall v \in L^1\big((0,T); BV(\Omega)\big) \cap L^2(\Omega_T), \end{aligned} \tag{11}$$

$$\begin{aligned} &\int_0^s \int_\Omega u^\varepsilon_{\rho t}(v - u^\varepsilon_\rho)\, dxdt \\ &+ \int_0^s \big[J_{\lambda,\varepsilon}(v) - J_{\lambda,\varepsilon}(u^\varepsilon_\rho)\big]\, dt \ge 0 \quad \forall v \in L^1\big((0,T); BV(\Omega)\big), \end{aligned} \tag{12}$$

and

$$\| u(s) - u_\rho(s) \|_{L^2} \le \| u_0 - (u_0)_\rho \|_{L^2} \qquad \forall s \in [0,T], \tag{13}$$

$$\| u^\varepsilon(s) - u^\varepsilon_\rho(s) \|_{L^2} \le \| u^\varepsilon_0 - (u^\varepsilon_0)_\rho \|_{L^2} \qquad \forall s \in [0,T]. \tag{14}$$

Choosing $v = u^\varepsilon_\rho$ in (11), $v = u_\rho$ in (11), and adding the resulting inequalities yield

$$\begin{aligned} &-\int_0^s \int_\Omega (u_\rho - u^\varepsilon_\rho)_t (u_\rho - u^\varepsilon_\rho)\, dx\, dt \\ &\qquad + \int_0^s \Big\{\big[J_\lambda(u^\varepsilon_\rho) - J_{\lambda,\varepsilon}(u^\varepsilon_\rho)\big] + \big[J_{\lambda,\varepsilon}(u_\rho) - J_\lambda(u_\rho)\big]\Big\}\, dt \ge 0, \end{aligned}$$

which and (9) imply that

$$\begin{aligned} \| u_\rho(s) - u^\varepsilon_\rho(s) \|^2_{L^2} &\le \| u_\rho(0) - u^\varepsilon_\rho(0) \|^2_{L^2} + 2\int_0^s \Big\{\big[J_\lambda(u^\varepsilon_\rho) - J_{\lambda,\varepsilon}(u^\varepsilon_\rho)\big] \\ &\qquad + \big[J_{\lambda,\varepsilon}(u_\rho) - J_\lambda(u_\rho)\big]\Big\}\, dt \\ &\le \| (u_0)_\rho - (u^\varepsilon_0)_\rho \|^2_{L^2} + 4C_0(T)\, \varepsilon^\alpha \qquad \forall s \in [0,T]. \end{aligned} \tag{15}$$

Finally, it follows from (13)–(15) and the triangle inequality that

$$\begin{aligned}\| u(s) - u^{\varepsilon}(s) \|_{L^2} &\le \| u(s) - u_{\rho}(s) \|_{L^2} + \| u_{\rho}(s) \\ &\qquad - u^{\varepsilon}_{\rho}(s) \|_{L^2} + \| u^{\varepsilon}_{\rho}(s) - u^{\varepsilon}(s) \|_{L^2} \\ &\le \| u_0 - (u_0)_{\rho} \|_{L^2} + \| (u_0)_{\rho} \\ &\qquad - (u^{\varepsilon}_0)_{\rho} \|_{L^2} + 2\sqrt{C_0(T)}\, \varepsilon^{\frac{\alpha}{2}} + \| u^{\varepsilon}_0 - (u^{\varepsilon}_0)_{\rho} \|_{L^2}.\end{aligned}$$

The desired estimate (10) follows from setting $\rho \to 0$ in the above inequality. The proof is complete. □

For readers' convenience, we now verify the assumption (9) for some commonly used regularization procedures.

Example 1 For any $1 < q < \infty$, define

$$J_{\lambda,\varepsilon,q}(u) := \int_{\Omega} f_{\varepsilon,q}(|Du|)\, dx + \frac{\lambda}{2} \int_{\Omega} (u - g)^2\, dx, \tag{16}$$

where $f_{\varepsilon,q}$ is given by

$$f'_{\varepsilon,q}(z) = \frac{z}{\sqrt[q]{z^q + \varepsilon^q}}.$$

Note that $f_{\varepsilon,2}(z) = f_{\varepsilon}(z) = \sqrt{z^2 + \varepsilon^2}$, this case was studied in great details in [15].

For any $u \in C^1(\Omega_T)$, a direct calculation yields

$$\begin{aligned}\big|J_{\lambda,\varepsilon,2}(u) - J_{\lambda}(u)\big| &= \int_{\Omega} \Big[\sqrt{|Du|^2 + \varepsilon^2} - |Du| \Big]\, dx \\ &= \int_{\Omega} \frac{\varepsilon^2}{\sqrt{|Du|^2 + \varepsilon^2} + |Du|} \le |\Omega|\, \varepsilon.\end{aligned} \tag{17}$$

Since $C^{\infty}(\Omega_T)$ is dense in $L^1((0,T); BV(\Omega)) \cap L^2(\Omega_T)$, it follows from above estimate and a standard density argumentation that (9) holds with $C_0(T) = |\Omega| T$, and $\alpha = 1$.

Remark 1 The above estimate gives the worst scenario. In the case that $|Du|$ has a positive lower bound almost everywhere, that is, $\{|Du| \ge c_0\} \approx \Omega$, we have $\alpha = 2$. Hence, we get linear rate of convergence in ε.

Example 2 The following regularization procedure is a modification of the one proposed and analyzed by Chambolle and Lions in [13]

$$J_{\lambda,\varepsilon,\mathrm{CL}}(u) := \int_{\Omega} \phi_{\varepsilon}(|Du|)\, dx + \frac{\lambda}{2} \int_{\Omega} (u - g)^2\, dx, \tag{18}$$

where ϕ_ε is given by

$$\phi_\varepsilon(z) := \begin{cases} \frac{1}{2\varepsilon} z^2 & \text{if } 0 \le z \le \varepsilon, \\ z - \frac{\varepsilon}{2} & \text{if } z \ge \varepsilon. \end{cases}$$

For any $u \in C^1(\Omega_T)$, a direct calculation gives

$$\begin{aligned} \left| J_{\lambda,\varepsilon,\mathrm{CL}}(u) - J_\lambda(u) \right| &= \left| \int_{\{|Du| \le \varepsilon\}} |Du| \left[\frac{1}{2\varepsilon} |Du| - 1 \right] dx - \int_{\{|Du| \ge \varepsilon\}} \frac{\varepsilon}{2} \, dx \right| \\ &\le |\Omega| \, \varepsilon + \frac{|\Omega|}{2} \, \varepsilon \le \frac{3}{2} |\Omega| \, \varepsilon, \end{aligned}$$

which and a standard density argumentation imply that (9) holds with $C_0(T) = \frac{3}{2}|\Omega| T$, and $\alpha = 1$.

3 Rate of convergence of finite element approximations

A fully discrete finite element method for the regularized flow (6)–(8) under certain numerical requirements was proposed in [15], and optimal order error estimates was established. In addition, it was shown that the finite element solution converges to the solution of the TV flow (2)–(4) as the mesh sizes and the parameter ε all tend to zero. On the other hand, no rate of convergence was given there.

In this section, we first propose and analyze a semi-discrete (in time) scheme to approximate the weak solution of the TV flow (2)–(4). In particular, we verify error estimates for the semi-discrete scheme. Then, we revisit the fully discrete finite element method developed in [15], and establish a rate of convergence for using the method to approximate the TV flow.

3.1 An implicit time discretization for the TV flow

Let $\{t_m\}_{m=0}^M$ be an equidistant partition of $[0, T]$ of mesh size $k \in (0, 1)$, and $d_t u^m := \frac{1}{k}(u^m - u^{m-1})$. Our semi-discrete in time scheme for approximating the TV flow (2)–(4) is defined as follows: Given $u^0 \in BV(\Omega) \cap L^2(\Omega)$, find $\{u^m\}_{m=1}^M \in BV(\Omega) \cap L^2(\Omega)$ such that

$$\begin{aligned} &\int_\Omega d_t u^m (w - u^m) \, dx + J_\lambda(w) - J_\lambda(u^m) \ge 0 \\ &\qquad \forall\, w \in BV(\Omega) \cap L^2(\Omega), \quad m = 1, 2 \cdots, M. \end{aligned} \tag{19}$$

Summing over m from 1 to ℓ ($1 \le \ell \le M$) after taking $w = u^{m-1}$ in (19) leads to the following a priori estimate for $\{u^m\}_{m=1}^M$

$$k \sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2 + J_\lambda(u^\ell) \le J_\lambda(u^0) \qquad \forall \ell \le M. \tag{20}$$

Well-posedness of problem (19) can be shown by following corresponding studies in [15]. Convergence behavior depends on regularity of initial data: if $u_0 \in \mathrm{Dom}(\partial J_\lambda)$, optimal order rate of convergence follows from a general result of Rulla [22] for the backward Euler time-discretization of differential inclusions $-u_t \in \mathcal{A}(u)$ with a maximal monotone operator $\mathcal{A} = \partial J_\lambda$. In case of more general initial data $u_0 \in \overline{\mathrm{Dom}(\partial J_\lambda)} \equiv L^2(\Omega)$, we prove a suboptimal order rate of convergence for the semi-discrete scheme.

Theorem 3 *(i) Suppose that $g \in L^2(\Omega)$ and $u_0 \in BV(\Omega) \cap L^2(\Omega)$. Let u, $\{u^m\}_{m=0}^M$ be the solutions of TV-flow and* (19)*, respectively. Define*

$$\overline{\overline{u}}^k(\cdot, t) := \frac{t - t_{m-1}}{k} u^m(\cdot) + \frac{t_m - t}{k} u^{m-1}(\cdot) \qquad \forall t \in [t_{m-1}, t_m].$$

Then, there holds

$$\operatorname*{ess\,sup}_{t\in[0,T]} \| u(t) - \overline{\overline{u}}^k(t) \|_{L^2(\Omega)} \le \| u_0 - u^0 \|_{L^2(\Omega)} + C\sqrt{k}\sqrt{J_\lambda(u_0) + J_\lambda(u^0)}. \tag{21}$$

(ii) If $u_0 \in Dom(\partial J_\lambda)$, there holds

$$\operatorname*{ess\,sup}_{t\in[0,T]} \| u(t) - \overline{\overline{u}}^k(t) \|_{L^2(\Omega)} \le \| u_0 - u^0 \|_{L^2(\Omega)} + Ck \, \| (\partial J_\lambda)^0(u_0) \|_{L^2} , \tag{22}$$

where $(\partial J_\lambda)^0(u_0)$ is the unique element of minimal norm from the closed and convex set $\partial J_\lambda(u_0)$.

Proof. Since (ii) follows from Theorem 5 of [22], it suffices to prove (i). Notice that (19) can be rewritten as

$$\int_\Omega \overline{\overline{u}}^k_t (w - \overline{u}^k)\, dx + J_\lambda(w) - J_\lambda(\overline{u}^k) \ge 0 \qquad \forall w \in BV(\Omega) \cap L^2(\Omega), \quad t \in (0, T), \tag{23}$$

where $\overline{u}^k(\cdot, t) := u^m(\cdot)$ for $t \in (t_{m-1}, t_m]$. Now, choosing $w = \overline{u}^k$ in a corresponding representation for TV-flow, with $s_0 = 0$ and $w = u$ in (23) we get

$$\int_\Omega u_t(\overline{u}^k - u)\, dx + J_\lambda(\overline{u}^k) - J_\lambda(u) \ge 0\,,$$
$$\int_\Omega \overline{\overline{u}}^k_t(u - \overline{u}^k)\, dx + J_\lambda(u) - J_\lambda(\overline{u}^k) \ge 0\,.$$

Adding the above two inequalities and using the notation

$$\overline{e}(\cdot, t) := u(\cdot, t) - \overline{u}^k(\cdot, t), \qquad \overline{\overline{e}}(\cdot, t) := u(\cdot, t) - \overline{\overline{u}}^k(\cdot, t) \qquad \forall t \in (t_{m-1}, t_m],$$

we get

$$\frac{1}{2}\frac{d}{dt}\| \overline{\overline{e}} \|_{L^2}^2 \le -\int_\Omega \overline{\overline{e}}_t(\overline{\overline{u}}^k - \overline{u}^k)\,dx\,. \tag{24}$$

It follows from a direct calculation and (20) that for any $1 \le \ell \le M$

$$\begin{aligned}\int_0^{t_\ell} \| \overline{u}^k - \overline{\overline{u}}^k \|_{L^2}^2\,dt &= \sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2 \int_{t_{m-1}}^{t_m} (t - t_{m-1})^2\,dt \\ &= \frac{k^3}{3}\sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2 \le C\,k^2,\end{aligned}$$

which, together with (20), then leads to

$$\begin{aligned}&\Big|\int_0^{t_\ell}\int_\Omega \overline{\overline{e}}_t(\overline{\overline{u}}^k - \overline{u}^k)\,dxdt\Big| \\ &\le \left\{\int_0^{t_\ell} \| u_t \|_{L^2}^2\,dt + k\sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2\right\}^{\frac{1}{2}} \left\{\int_0^{t_\ell} \| \overline{u}^k - \overline{\overline{u}}^k \|_{L^2}^2\,dt\right\}^{\frac{1}{2}} \\ &\le C\,k. \end{aligned} \tag{25}$$

Finally, the desired estimate (21) follows from integrating (24) from 0 to t_ℓ and appealing to (9) and (25). The proof is complete. □

Efficient control of spatial discretization effects requires additional regularity properties of solutions which is another motivation for us to come back to a fully discrete version of the discretization (6)–(8) in the next subsection.

3.2 A fully discrete finite element method for the TV flow

Let $\mathcal{T}_h$ be a quasiuniform triangulation of Ω with mesh size $h \in (0, 1)$, and V^h denote the continuous, piecewise linear finite element space associated with $\mathcal{T}_h$, that is,

$$V^h := \left\{ v_h \in C^0(\overline{\Omega});\ v_h\big|_K \in P_1(K),\ \forall\, K \in \mathcal{T}_h \right\}.$$

We recall that the fully discrete finite element method of [15] for the gradient flow (6)–(8) is defined as follows: Given $U^0 \in V^h$, find $\{U^m\}_{m=1}^M \in V^h$ such that

$$\big(d_t U^m, v_h\big) + \Big(\frac{f_\varepsilon'(|\nabla U^m|)}{|\nabla U^m|}\nabla U^m, \nabla v_h\Big) + \lambda\big(U^m - g, v_h\big) = 0 \qquad \forall\, v_h \in V^h, \tag{26}$$

where $(\cdot,\cdot)$ stands for the standard $L^2(\Omega)$ inner product, and $f_\varepsilon(z) = \sqrt{z^2 + \varepsilon^2}$.

Define the linear interpolation of $\{U^m\}_{m=0}^M$ in time as

$$\overline{\overline{U}}^{\varepsilon,h,k}(\cdot,t) := \frac{t-t_{m-1}}{k}U^m(\cdot) + \frac{t_m - t}{k}U^{m-1}(\cdot) \qquad \forall\, t \in [t_{m-1}, t_m]\,. \tag{27}$$

If $u_0 \in W^{1,1}(\Omega) \cap H^1_{loc}(\Omega)$ and $g \in L^2(\Omega) \cap H^1_{loc}(\Omega)$, it was shown in Theorem 1.6 of [15] that

$$\lim_{\varepsilon\to 0}\lim_{h,k\to 0} \| u - \overline{\overline{U}}^{\varepsilon,h,k} \|_{L^\infty(0,T);L^p(\Omega))} = 0\,, \qquad 1 \le p < \frac{d}{d-1}\,, \tag{28}$$

provided that $\lim_{h\to 0} \| u_0 - U^0 \|_{L^2} = 0$.

We now state a theorem which provides a rate of convergence for the error $u - \overline{\overline{U}}^{\varepsilon,h,k}$.

Theorem 4 *In addition to the assumptions of Theorem 3, suppose that* $u_0 \in C^2(\overline{\Omega})$, $g \in L^\infty\big((0,T); W^{1,\infty}(\Omega)\big)$, $\partial\Omega \in C^3$, *and* $f_\varepsilon(z) = \sqrt{z^2+\varepsilon^2}$. *Then, under the following starting value and mesh constraints*

$$\| u_0^\varepsilon - U^0 \|_{L^2} \le Ch^2 \qquad \text{and} \qquad k = \mathcal{O}(h^2),$$

there holds the error estimate

$$\begin{aligned} \operatorname*{ess\,sup}_{t\in[0,T]} \| u(t) - \overline{\overline{U}}^{\varepsilon,h,k}(t) \|_{L^2(\Omega)} &\le \| u_0 - u_0^\varepsilon \|_{L^2} + 2\sqrt{|\Omega| T}\,\sqrt{\varepsilon} \\ &\quad + C_1(\varepsilon)k + C_2(\varepsilon)h^2\,, \end{aligned} \tag{29}$$

where $C_i(\varepsilon)$ *for* $i = 1, 2$ *are positive constants which depend on* ε^{-1} *in some low polynomial order.*

Proof. (29) follows immediately from the triangle inequality

$$\| u - \overline{\overline{U}}^{\varepsilon,h,k} \|_{L^2} \le \| u - u^\varepsilon \|_{L^2} + \| u^\varepsilon - \overline{\overline{U}}^{\varepsilon,h,k} \|_{L^2},$$

and appealing to (10) and Theorem 1.7 of [15]. The proof is complete. □

Remark 2 (a). Assumptions in Theorem 4 require data (u_0, g) to be regular, and (29) suggests that any (smooth) approximation $(\tilde{u}_0, \tilde{g})$ of (u_0, g) satisfying

$$\| u_0 - \tilde{u}_0 \|_{L^2(\Omega)} + \| g - \tilde{g} \|_{L^2(\Omega)} = \mathcal{O}(\sqrt{\varepsilon})$$

should be sufficient.

(b). The coefficients $C_1(\varepsilon)$ and $C_2(\varepsilon)$ in (29) depend on ε^{-1} in some low polynomial orders, and no sharp bounds are available from the analysis. In Section 4, we will provide computational evidence for appropriate scalings for k and h with respect to ε in order to obtain optimal order convergence.

4 Numerical experiments

We present some numerical experiments for the regularization procedures (16) and (18), and numerically identify scaling laws between mesh sizes h, k and regularization parameter ε. In all our numerical tests, we solve the nonlinear (algebraic) equations at each time step using a fixed-point iteration, where the diffusivity is evaluated at the previous iterate. This algorithm performs remarkably well, for all regularizations we obtained good approximations after only few iterations. Furthermore, the number of iterations required to trigger the stopping criteria is essentially independent of the regularization procedures. In that sense, we found that all regularization strategies lead to efficient numerical schemes of the same quality.

Figure 4 (a) displays L^2-errors at $t = 5 \cdot 10^{-3}$ for $f_\varepsilon = f_{\varepsilon,2}$ for different scalings $\varepsilon = \mathcal{O}(k^r)$ for $r = 0, 1/2, 1, 3/2$, which validates a proper scaling law $\varepsilon = \mathcal{O}(k)$; this scaling law is better compared to the theoretically predicted one in Theorem 4, which may be explained by the 'pessimistic' bound in (17) in situations when $\{|\,Du\,| > c\}$ almost covers Ω. Figure 5 displays

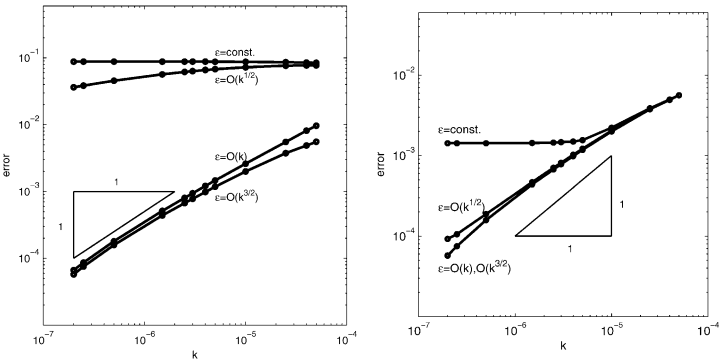

Fig. 4. L^2-error at $t = 5 \cdot 10^{-3}$ with different scaling laws $\varepsilon = \mathcal{O}(k^r)$, using regularization (a) $f_{\varepsilon,2}$, (b) $f_{\varepsilon,\mathrm{CL}}$

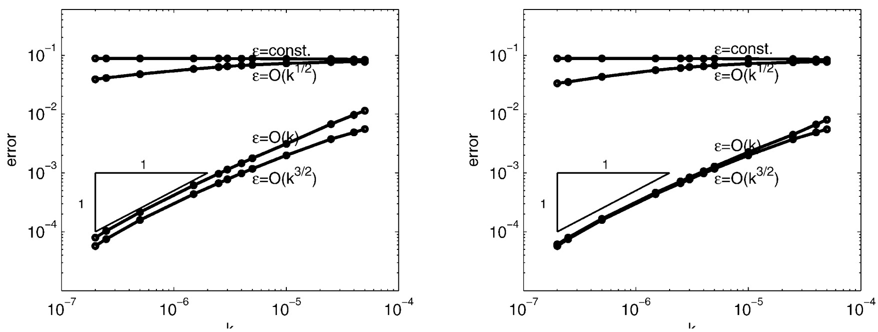

Fig. 5. L^2-error at $t = 5 \cdot 10^{-3}$ using $f_{\varepsilon,q}$-regularization with different scaling laws $\varepsilon = \mathcal{O}(k^r)$. $q = 1.5$ (left), $q = 5$ (right)

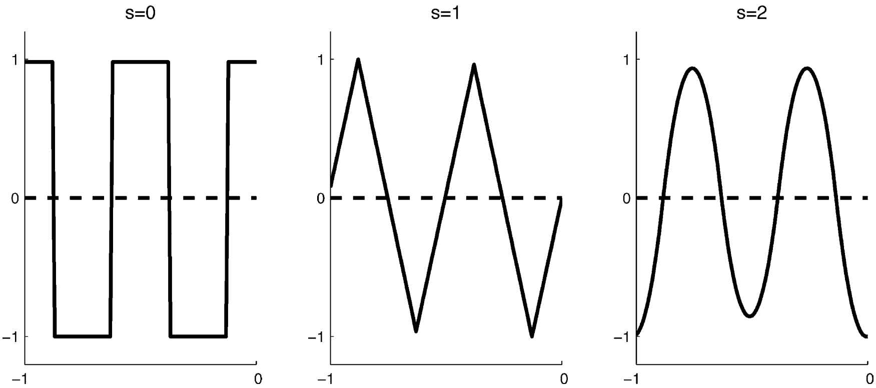

Fig. 6. Piecewise polynomial initial data with discontinuities in the s-th derivative for $s = 0, 1, 2$

corresponding results for the regularizations $f_{\varepsilon,q}$, for $q = 1.5$ and 5; Analogous experiments for the case $f_{\varepsilon,\text{CL}}$-regularization are shown in Figure 4 (b), which again evidence an optimal scaling $\varepsilon = \mathcal{O}(k)$.

Spatial discretization effect in the case of $f_{\varepsilon,2}$-regularization is reported in Figure 7 using initial data of different regularities (see Figure 6). We observe a decrease in rate of convergence when using rough initial data (cf. Theorem 4).

Finally, we examine the regularization procedures using $f_{\varepsilon,\text{CL}}$ and $f_{\varepsilon,q}$ for $q = 1, 2, 100$. Figure 8 plots the diffusivity function $f'_{\varepsilon,*}(s)/s$ of each regularization. We remark that the regularization procedure of Chambolle and Lions [13] amounts to applying the heat equation (constant diffusivity) in regions of small gradients and using the TV flow at places where the gradient

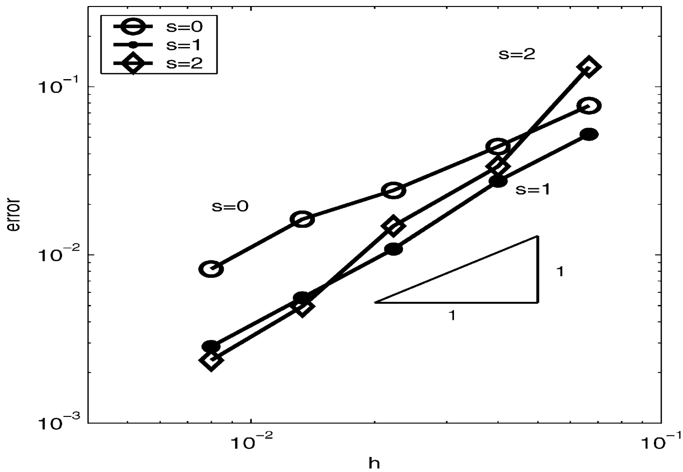

Fig. 7. L^2-error at $t = 10^{-3}$ using $f_{\varepsilon,2}$-regularization. $\varepsilon = 10^{-3}$, $k = 10^{-6}$, and the initial data of Figure 6 are used in the tests

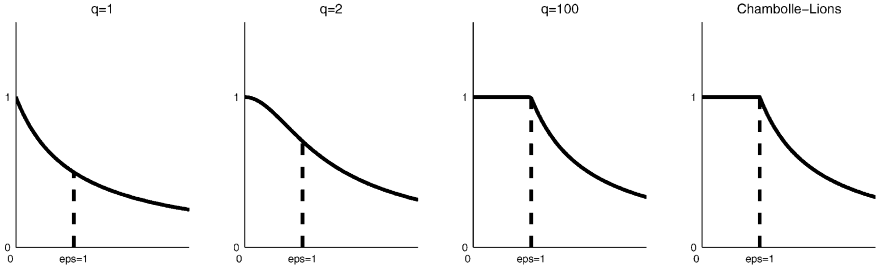

Fig. 8. The diffusivity for $f_{\varepsilon,q}$, $q = 1, 2, 100$ and for $f_{\varepsilon,\mathrm{CL}}$

is big. For large q, the $f_{\varepsilon,q}$-regularization has similar properties. However, for small q the diffusivity of $f_{\varepsilon,q}$ is large for small gradients and small for large gradients where the edges reside. For two-dimensional image denoising applications, our numerical tests indicate that all these regularization strategies essentially perform equally well. Figure 3 depicts three snapshots for the regularized flow with $f_{\varepsilon,2}$. Corresponding tests with the Chambolle-Lions regularization $f_{\varepsilon,\mathrm{CL}}$ and the regularizations $f_{\varepsilon,q}$ for $q \neq 2$ produce results which do not differ qualitatively from this study.

Acknowledgements. The first and the third authors would like to thank the Mathematisches Forschungsinstitut Oberwolfach for the kind hospitality and opportunity of its "Research in Pairs" program. All three authors would like to thank J. Vogelgesang (ETH Zürich) for providing them with Fig. 1 and Fig. 2.

References

1. Acar, R., Vogel, C.R.: Analysis of bounded penalty methods for ill-posed problems. Inv. Problems **10**, 1217–1229 (1994)
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. Oxford University Press, New York 2000
3. Andreu, F., Ballester, C., Caselles, V., Mazón, J.M.: The Dirichlet problem for the total variation flow. J. Funct. Anal. **180**, 347–403 (2001)
4. Andreu, F., Ballester, C., Caselles, V., Mazón, J.M.: Minimizing total variation flow. Diff. Int. Equations **14**, 321–360 (2001)
5. Andreu-Vaillo, F., Caselles, V., Mazón, J.M.: Parabolic quasilinear equations minimizing linear growth functionals. Birkhäuser, 2004
6. Andreu, F., Caselles, V., Díaz, J.I., Mazón, J.M.: Some qualitative properties for the total variation flow. J. Funct. Anal. **188**, 516–547 (2002)
7. Anzellotti, G.: Pairings between measures and bounded functions and compensated compactness. Ann. di Matematica Pura ed Appl. IV **135**, 293–318 (1983)
8. Bellettini, G., Caselles, V., Novaga, M.: The total variation flow in $\mathbf{R}^N$. J. Diff. Eqns. **184**, 475–525 (2002)
9. Brezis, H.: Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland, 1973
10. Bruck, R.E.: Asymptotic convergence of nonlinear contraction semi-groups in Hilbert spaces. J. Funct. Anal. **18**, 15–26 (1975)

11. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vision **20**, 89–97 (2004)
12. Chan, T., Shen, J.: On the role of the BV image model in image restoration. Tech. Report CAM 02-14, Department of Mathematics, UCLA, 2002
13. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**, 167–188 (1997)
14. Dobson, D.C., Vogel, C.R.: Convergence of an iterative method for total variation denoising. SIAM J. Numer. Anal. **34**, 1779–1791 (1997)
15. Feng, X., Prohl, A.: Analysis of total variation flow and its finite element approximations. Math. Mod. Num. Analysis **37**, 533–556 (2003)
16. Feng, X., Oehsen, M.V., Prohl, A.: Rate of convergence of regularization procedures and finite element approximations for the total variation flow, ETH-Report **12**, (2003)
17. Gerhardt, C.: Evolutionary surfaces of prescribed mean curvature. J. Diff. Equations **36**, 139–172 (1980)
18. Hardt, R., Zhou, X.: An evolution problem for linear growth functionals. Commun. PDEs **19**, 1879–1907 (1994)
19. Kobayashi, R., Warren, J.A., Carter, W.C.: A simple continuum model of grain boundaries. Physica D, **140**, 141–150 (2000)
20. Lichnewsky, A., Temam, R.: Pseudosolutions of the time-dependent minimal surface problem. J. Diff. Equations **30**, 340–364 (1978)
21. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
22. Rulla, J.: Error analysis for implicit approximations to solutions to Cauchy problems. SIAM J. Num. Analysis **33**, 68–87 (1996)
23. Vese, L.: A study in the BV Space of a denoising-deblurring variational problem. Appl. Math. & Opt. **44**, 131–161 (2001)